Association for Information Systems

AIS Electronic Library (AISeL)

BLED 2019 Proceedings

BLED Proceedings

2019

Personality-Based Content Engineering for Rich Digital Media

Haris Krijestorac,

Rajiv Garg

Maytal Saar-Tsechansky

32ND BLED ECONFERENCE
HUMANIZING TECHNOLOGY FOR A SUSTAINABLE SOCIETY
JUNE 16 – 19, 2019, BLED, SLOVENIA, CONFERENCE PROCEEDINGS
A. Pucihar (ed. et al.)

Personality-Based Content Engineering for Rich Digital Media

HARIS KRIJESTORAC, RAJIV GARG & MAYTAL SAAR-TSECHANSKY

Abstract Firms have increasingly turned to rich digital media, such as videos and photos, to attract attention and boost awareness. Although extant research may help firms promote these media more effectively, the marketing process truly begins with creation of the media. Thus, content creators may benefit from understanding what media is likely to achieve greater popularity, based on its content features. We develop a method to understand the effect of content on the consumption of online videos, and employ our method on a unique dataset including 16,414 videos from 363 YouTube channels. Our approach labels videos as high- or low-performing relative to comparable videos, and leverages random forests to identify content features associated with performance level. We test this method using the personality of speech-driven videos, employing NLP to estimate the extent to which video captions exhibit each of the "big five" personality traits. Our analysis uncovers predictive, economic, and prescriptive insights. We find that using just their personality, we can predict whether videos perform better than expectation with 72% accuracy. Furthermore, videos associated with high-performing personalities can expect a nearly 15% increase in consumption. Finally, we examine which personalities are associated with high consumption, offering prescriptive insights for content engineering.

Keywords: • Content engineering • Personality • Rich digital media• random forests • NLP •

CORRESPONDENCE ADDRESS: Haris Krijestorac, PhD Student, University of Texas at Austin, McCombs School of Business, Texas, United States of America, e-mail: haris.krijestorac@mccombs.utexas.edu. Rajiv Garg, PhD, Assistant Professor, University of Texas at Austin, McCombs School of Business, Texas, United States of America, e-mail: rajiv.garg@mccombs.utexas.edu. Maytal Saar-Tsechansky, PhD, Professor, University of Texas at Austin, McCombs School of Business, Texas, United States of America, e-mail: maytal.saar-tsechansky@mccombs.utexas.edu.

DOI https://doi.org/10.18690/978-961-286-280-0.48 ISBN 978-961-286-280-0
Dostopno na: http://press.um.si

1 Introduction

As the Internet enhances users' ability to filter and search for information, it is imperative that firms "pull" in consumers using content marketing, rather than merely "pushing" their message through paid advertising. The effectiveness of "push" strategies has been recently questioned, due to consumer resistance to advertising (Baek and Morimoto 2013)[1] and rising adblocker adoption[2]. In response, firms have turned to rich digital media, such as videos, images, and whitepapers, to promote themselves online (Corcoran 2009). In doing so, firms can leverage their fan base to spread word-of-mout (WOM) about their media, which may boost brand perceptions before the point of sale (Järvinen and Taiminen 2016).

But although content marketing offers promotional opportunities for firms, it also presents unique challenges. While firms can use analytical tools to iteratively refine the targeting, budget, and messaging associated with their ads, digital media offer less opportunity for modification after the media are released. As such, the content design of these media is a key determinant of their success. Presently, the creation of digital media is largely driven by creativity and intuition, and often lacks empirical guidance. This is perhaps due to the perception that the spread of digital media is somewhat random and unmanageable (Bampo 2008), or that features of these media that lead to their popularity are hard to quantify or operationalize. In this paper, we will provide evidence that, in fact, the content features of digital media *can* predict their consumption, and that high-performing features can be *learned* through empirical analysis. We will introduce and test an approach to identifying these high-performing content features, and estimate the effectiveness of these features in boosting consumption of media.

While advances in machine learning (e.g., deep learning, NLP) present opportunities to capture the content features associated with digital media, it is unclear how to assess the effect of these features on media consumption. Thus, it remains uncertain whether the content of these media can inform more effective media design, and if so, how firms can identify content features that are

[1] For example, organic results on Google searches achieve over twenty times the clicks of paid results, despite occupying less than forty percent of screen real estate (source: https://sparktoro.com/blog/seo-opportunity-growing-shrinking/).
[2] https://www.emarketer.com/Article/Ad-Blocker-Use-Grows-Publishers-Face-New-Challenges/1016076

associated with high-performing media. In addition to uncovering effective content features, firms and content creators could benefit from understanding what rewards they can expect to reap by creating media that reflect these features. To address these issues, our study poses two questions: (1) How can the role of content features in consumption of media be learned, and (2) What increase in consumption is associated with media that exhibits high-performing features?

Our analysis of content of digital media focuses on the "Big Five" personality traits (Norman 1963). While our model can be extended to incorporate other content features (e.g., tone, visual elements), we focus on personality due to its demonstrated relevance to the spread of WOM (Adamopoulos et al. 2018; Devaraj et al. 2008; McElroy et al. 2007), which is a key driver of consumption for digital media (Susarla et al. 2011). The "Big Five" traits refer to a psycholinguistic framework which advocates that personality can be characterized by five dimensions: openness, conscientiousness, extroversion, agreeableness, and neuroticism. Individuals or collectives may exhibit varying degrees of each of these traits, which can be described as follows. The *openness* trait is associated with curiosity and willingness to try new experiences, while individuals with low openness tend to be more cautious and reserved. *Conscientiousness* refers to a preference for planning over spontaneity, and self-discipline over free-spiritedness. The trait of *extroversion* is marked by high levels of engagement with other individuals and the external world, whereas individuals who are less extroverted (i.e., introverted) are more introspective. *Agreeableness* suggests an attentiveness towards others, and a concern for social and interpersonal harmony. Finally, the *neuroticism* trait is associated with emotional volatility, and a tendency to get irritated easily. The degree to which this collection of traits is exhibited by an entity can be referred to as its "personality".

The extent to which each personality trait is exhibited can be inferred through behaviors or speech (Corr and Matthews 2009; Costa & McCrae 1992). Although personality may vary slightly over time or adapt to situational factors, there is evidence that personalities tend to be relatively stable, and even biologically-influenced (Briley and Tucker-Drob 2014). As such, personality has been shown to various outcomes, including brand evangelism (Doss & Karstens 2014), job performance (Barrick and Mount 1991), and marital stability (Kelly & Conley 1987).

While much of the psychological literature has considered outcomes associated with the personality of "real" individuals or collectives, personality can also be exhibited by entities created with human input (Aaker 1997; Corr and Matthews 2009). Accordingly, we propose that the characters, personas, or narrators (whether real or fictitious) within media may project discernable personalities. These personalities may reflect those of real individuals within the media, or of individuals who were involved in creating the media, such as a firm's employees or management. Firms thus make (at least implicit) decisions about what personalities to project in their media; Our aim is therefore to investigate whether firms can uncover empirical insights to understand what personalities may be more effective at appealing to their audience, and thus generate greater consumption.

In assessing the role of personality on the consumption of online videos, we contribute to a nascent, yet growing body of literature on *content engineering*. This research explores the role of content features on the effectiveness of various media, including traditional offline ads, digital ads, and non-advertising media. For example, research has examined the effects of appealing to intuition vs. reason in direct advertising (Bertrand et al. 2010), and of focusing on action, information, or emotion in TV advertising (Liaukonyte et al. 2015). In a digital context, research has also considered the benefits of ad personalization (Tucker 2014), and of features of Facebook posts on user engagement (Lee et al. 2018). Finally, some work has explored the role of content for media besides advertisements, including the effect of emotional content on the digital sharing of *New York Times* articles (Berger and Milkman 2012), and the effect of political slant on newspaper readership (Gentzkow and Shapiro 2010).

While prior research suggests an association between content and the spread of WOM, we build on that work in three key ways. First, previous studies focus on media whose popularity is inherently ephemeral, such as newspapers providing timely information, or Facebook posts that quickly get drowned out in new content. However, our study focuses on online videos, which exhibit dynamic consumption patterns that we can examine over an extended time period. Second, this study is the first, to our knowledge, to examine the role of personality in digital media consumption. In doing so, we answer calls to consider the role of rich, multi-dimensional aspects of human behavior, and moving beyond binary or one-dimensional metrics such as sentiment (Kim et al. 2013).

Given increasing sophistication of content analysis techniques, our analyses may provide a useful example to inspire future research. Third, our study uncovers not only "global" insights that may apply to all content creators, but also demonstrates that it can be even more useful to identify high-performing content features for a particular content creator.

2 Theory and Methdology

2.1 How Content Personality Affects Media Consumption

To understand the connection between personality and media consumption, we draw from literature on WOM, information diffusion, and personality theory to propose the mechanism illustrated in Figure 1. First, we note that any speech-driven media featuring a focal persona or narrator will project their personality through its speech. Although personality is considered an internal property of individuals, it can be inferred with high accuracy by observing external constructs such as word choice and phrasing (Fast and Funder 2008; Goldbeck et al. 2011). Moreover, speech can be used to infer personalities associated not only with individuals (Roccas et al. 2002; Salgado 2003), but also with entities created with human input (Aaker 1997; Corr and Matthews 2009), including digital media (Nass and Lee 2001). Thus, we can reliably extract personality traits from the speech associated with a video's narrator, without directly administering a psychological personality test (Hirsch and Peterson 2009). Although personality may correlate with features besides speech, text-based methods are considered the best and most accurate way to infer personality (Pennebaker 2013, ch. 4). Hence, one can argue that through the script or speech associated with a video's protagonist, we can infer the personality of the media. It is worth noting that we focus on this overall personality, rather than the words themselves, because higher-level constructs such as personality will offer more prescriptive value (Song et al. 2013).

Given that speech-driven media exhibit personalities, we consider how this personality could affect media consumption. We argue that the personality of media can affect the nature of the WOM around the media. Depending on who the content of the media appeals to, the quantity and quality of this WOM may vary. One reason for this may be that the media content resonates with its consumers by exhibiting similar personalities. Because interpersonal similarity tends to increase trust and decrease social friction (Byrne 1997; Lichetenthal and Tellefsen 2001), one can argue that media with personalities that match those of their consumer base may tend to generate more WOM. Similar benefits of personality similarity have been found in a variety of contexts, including friendship formation (Morry 2007), marital stability (Kelly & Conley 1987), and attraction to employers by employees (Devendorf and Highhouse 2010). In addition, individuals may be more prone to sharing media that reflects a personality to which they aspire (Malhotra 1988; Sirgy 1992). For example, if individuals wish to be perceived as more open, they may be more drawn to other individuals (or media) that exhibit high openness, regardless of whether they exhibit this trait themselves. Thus, due to interpersonal similarity and/or personal aspirations, the personality of media content may affect the personality of those who propagate WOM around the media upon consuming it. In turn, the personality of these consumers may affect the quantity, quality, and effectiveness of the WOM spread around the media (Adamopoulos et al. 2018; Devaraj et al. 2008; Lee et al. 2014).

Because WOM is a key driver of the popularity of digital media (Susarla et al. 2011), it is likely that the nature of this WOM will, in turn, affect the consumption of the media. As more individuals consume these media based on online chatter, they too may spread WOM around the media, creating a positive feedback loop between WOM and video consumption.

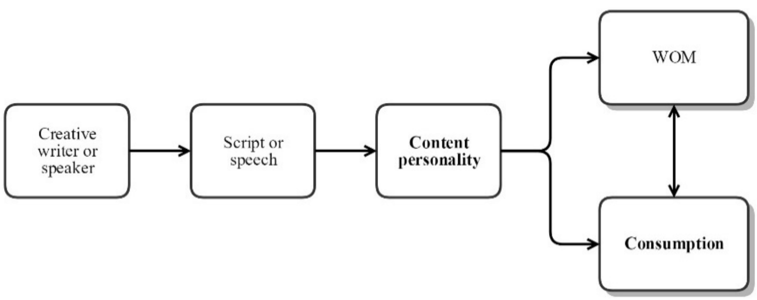

Figure 1. How Media Content Affects Consumption

2.2 Measuring Effect of Media Consumption on Personality

To extract predictive, economic, and prescriptive insights into the relationship between media content and its consumption, we introduce a three-step approach, outlined in Figure 2. Although we employ this approach in the context of online videos and personality, the methodology would allow for consideration of other content features (e.g., visuals, tone) as appropriate for a given context, as well as other media formats (e.g., images).

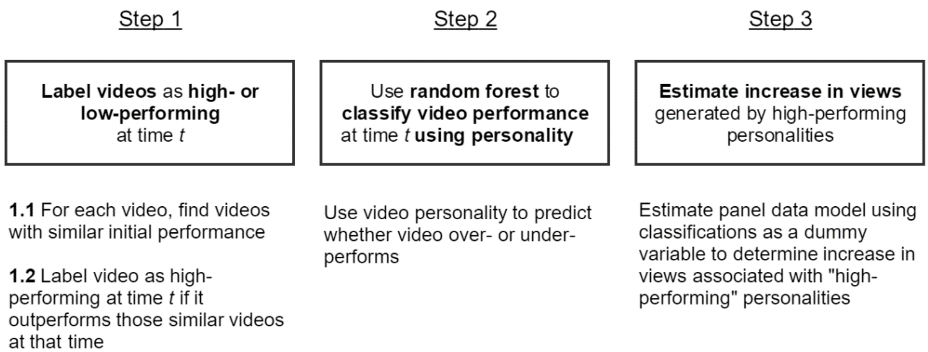

Figure 2. Overview of Methodology

To establish whether personality affects the consumption of online videos, we would first like to understand the predictive role of personality. To achieve this, our first step is to label each video based on the question: given its initial consumption momentum, does the video over- or under-perform? Because the early consumption of digital media reasonably predicts its long-term success (Szabo and Huberman 2010; Pinto et al. 2013)[3], one can argue that videos with similar initial views have similar potential. Per the "She's Mercedes" example in Figure 2, the early performance of media may be influenced by their superficial characteristics (e.g., brand, followers, audience size). On the other hand, as media are consumed over time, WOM plays an increasing role in their diffusion (Mahajan and Peterson 1985), suggesting that the actual content of the media will play a more salient role in their long-term consumption. Hence, our approach focuses on categorizing videos by performance relative to expectation, so that

[3] Past performance of a video is regarded as the best predictor of its future performance, regardless of the content or channel associated with the video.

we can eventually identify content features associated with over- and under-performance.

To estimate a video's expected views on a specific day *t*, we take the average of views on day *t* of a set of videos that had similar consumption to the focal video in their first 24 hours (i.e., on day one). We label videos that outperform their corresponding similar videos as "high performing" (i.e., labeled as "1") on day *t*, whereas those that underperform relative to expectation are labeled as "low-performing" (i.e., labeled "0"). We achieve this labelling in a two-step process, characterized by equations 1 and 2 below. Using Equation 1, we calculate the similarity in initial views of a pair of videos by simply taking the absolute value of the difference between these two view counts. In Equation 2, we obtain the median number of views on day *t* of *k* videos with the most similar initial performance. For our main analysis, we will use twenty videos with the most similar initial performance to a focal video (i.e., *k* = 20) – we will assess the robustness of this choice later. Finally, in Equation 3, we give video *i* label $H_i(t)$ in accordance with its relative performance at time *t*.

Equation 1. **Similarity in initial performance between video *i* and video *i'***

$$s(v_i, v_i') = \sqrt{(Y_i(1) - Y_i'(1))^2}$$

Equation 2. **Expected Views of video *i* at time *t***

$$E[Y_i(\mathrm{t})] = \mathrm{median}(Y_i^k(t))$$

Equation 3. **Labelling Videos by Performance**

$$Y_i(t) > E[Y_i(\mathrm{t})] \rightarrow H_i(t) = 1$$
$$Y_i(t) < E[Y_i(\mathrm{t})] \rightarrow H_i(t) = 0$$

Where $s(v_i, v'_i)$ = similarity in views between video *i* and video *i'* on day 1
k = the number of videos similar to *i* used to calculate expected future performance
$Y_i^k(t)$ = views of the *k*th most similar video to video *i* on day *t*

Using these performance labels, we would like to assess the role of personality on media consumption. To do this, we use random forests to estimate the effect of video personalities (as independent variables), on the video labels (as the dependent variable). Hence, these forests classify video performance using only personality traits. It is worth noting that labels associated with a single video may evolve over time. To account for account for a potentially evolving role of personality in the consumption of videos, we estimate random forests at each time.

Since random forest prediction accuracy may depend on the underlying heterogeneity of training data, we test our approach using different training sets – namely, channel-specific (somewhat heterogeneous) and global (highly heterogeneous). While global forests allow us to uncover overall insights into what personalities are high-performing, channel-specific forests will identify personalities that are high-performing for a given content creator. The global forests use all videos in our dataset to learn what personalities are associated with high- and low-performing videos. Meanwhile, channel-specific forests use only videos within a given channel to predict performance based on personality[4]. Intuitively, a given channel may have specific personalities that perform well for their audience. For example, a channel such as McDonalds may cater to a different audience than that of Business Insider, and thus may benefit from different personalities being reflected in their media. These differences in personality may be due to factors such as differing audience demographics, or differing products offered by each firm. Furthermore, even if we can identify high-performing personalities for specific channels, it does not necessarily follow that there would be personalities that tend to perform well across all content creators. Thus, both global and channel-specific models warrant investigation.

Given the assessment of the predictive role of personality in video popularity, the next part of our analysis measures the economic effect of personality on views. That is, if we classify a video (whether accurately or not) as high-performing, what change in views can the video expect to see? To quantify this effect, we estimate a panel data model that considers the role of relevant parameters on the daily views achieved by videos. To estimate the expected benefits of a video being classified as high-performing, we include this personality-based performance classification as an independent variable in our model. To extract further insights on the role of personality traits, we additionally

[4] We estimate channel-specific forests for 80 channels for which we have data on at least 30 videos associated with the channel.

control for percentile scores of each "big five" trait. To isolate the effect of personality from time-specific shocks in WOM that may affect views, we consider the change in number of likes and dislikes in each time period. As a robustness check, we estimate this model with not only the video's classification, but with its likelihood score produced during the random forest estimation. This score reflects the proportion of trees that classified the video as "high-performing", and thus represents the confidence level associated with the classification. We estimate the below model using both global and channel-specific classifications.

Model 1. The Effect of Personality Classification on Video Views

$$\log Y_{ij,t} - logY_{ij,t-1} = \beta_0 + \beta_1 * \boldsymbol{C_{ij,t}} + \beta_P * P_{ij} + \beta_2 * \Delta likes_{ij,t} + \beta_3 * \Delta dislikes_{ij,t} + \varepsilon_{i,t}$$

$$\log Y_{ij,t} - logY_{ij,t-1} = \beta_0 + \beta_1 * \boldsymbol{L_{ij,t}} + \beta_P * P_{ij} + \beta_2 * \Delta likes_{ij,t} + \beta_3 * \Delta dislikes_{ij,t} + \varepsilon_{i,t}$$

Where $Y_{ij,t}$ = daily views for video *i* in channel *j* at time *t*

$C_{ij,t}$ = Classification (0 or 1) associated with video *i* in channel *j* at time *t*

$L_{ij,t}$ = Proportion of trees classifying video *i* in channel *j* as high-performing at time *t*

P_{ij} = Vector of percentile scores associated with each "big five" personality traits associated with video *i* in channel *j*

$\Delta likes_{ij,t}$ = change in number of likes associated with video *i* in channel *j* between time *t* and *t-1*

$\Delta likes_{ij,t}$ = change in number of dislikes associated with video *i* in channel *j* between time *t* and *t-1*

With our understanding of the predictive and economic effects of personality on views, a logical next step would be to examine what personalities are most effective in increasing views. To do so, we will cluster videos by personality, and examine the average performance classifications of each personality cluster.

3 Data

We examine the role of content personality on the consumption of digital media using a unique dataset consisting of 16,414 online videos from 363 YouTube channels. We collected daily statistics on these videos and channels over 365 days, as represented in Figure 3 and summarized in Table 1 below.

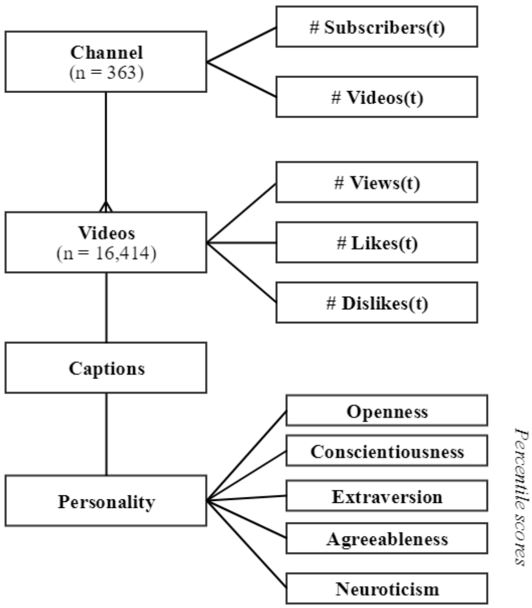

Figure 3. Data Overview

Table 1. Summary Statistics of Video and Channel Attributes

Attribute		Mean (St Dev)
Video	Caption length (# words)	923 (1,172)
Channel	# subscribers (max.)	2,595,446 (5,662,767)
Channel	# videos (max.)	1,760.91 (7,122.04)

Because we want to assess the relationship between videos' consumption patterns their personality, our data collection approach selects videos that were more likely to exhibit a discernable personality. To this end, we targeted videos that were speech-driven, rather than oriented around musical or visual elements. To identify such videos, we created a script that determined whether a video contained caption text, and whether this caption text was uploaded by the channel itself (as opposed to auto-generated)[5]. We target videos that contain caption text because psycholinguistic theory suggests that personality can be best inferred through speech, while theory that may help us infer personality through other components (e.g., visuals, music) is much less established. Speech-driven videos are therefore appropriate to the context of our study, as these videos typically have a discernable narrator or protagonist, who will project a specific personality.

To identify channels containing videos satisfying the aforementioned criteria, we considered channels associated with Fortune 500 companies, as well as over 1,000 channels associated with trending videos on YouTube over a two-week period[6]. We identified 61 such channels from the Fortune 500, and 302 channels associated with YouTube trending videos. To collect daily data on these videos from 363 channels, we executed a daily script. We obtained overvations on 6,640 videos with over 100 words of caption data, along with daily view, like, and dislike statistics over 365 days.

To infer the personalities of videos using caption text, we used IBM Watson Personality Insights[7]. Watson estimates the personality of a text corpus using natural language processing (NLP). This NLP analysis leverages a large database of diverse text corpuses with pre-labeled (i.e., known) personalities, and thus estimates the personality of each new corpus in a supervised manner. Based on this semantic similarity between video caption text and these pre-labeled corpuses, the IBM service outputs a vector with percentile scores for each of the "Big Five" personality traits.

Summary statistics on personality scores in Table 2 suggest that, with regard to most traits, YouTube videos are representative of personalities among the population. The one exception is openness, which is exhibited to a greater extent on YouTube captions than in other text corpuses.

[5] Captions can be viewed on YouTube by clicking on the 'CC' button on the lower-right corner of the video.
[6] The videos that were actually trending were not included in our dataset. Rather, trending videos helped us identify channels from which we would collect future data.
[7] https://console.bluemix.net/docs/services/personality-insights/science.html#science

Table 2. Summary Statistics for Video Personality Trait Percentiles

Personality Trait	Mean (St Dev)
Openness	.90 (.17)
Conscientiousness	.51 (.27)
Extroversion	.41 (.29)
Agreeableness	.33 (.33)
Neuroticism	.65 (.25)

While our data collection process selects a targeted sample of vidoes, our analysis approach can be implemented using various video features (e.g., visuals, tone). Speech-driven videos are simply selected due to the focus on personality.

4 Solution Design and Development

4.1 Predictive Insights: Personality and Video Performance

Per the first step of our methodology, we label all videos as high- or low-performing, based on whether they have more of fewer views than the median number of views of twenty similar videos at a particular time *t*. Using these labels, we estimate global and channel-specific random forests that classify video performance at each time period, using only personality traits as independent variables. Observing the accuracy of our random forest classifications in Figure 4, we find that both global and channel-level estimations achieve an accuracy greater than 50% over all time periods.

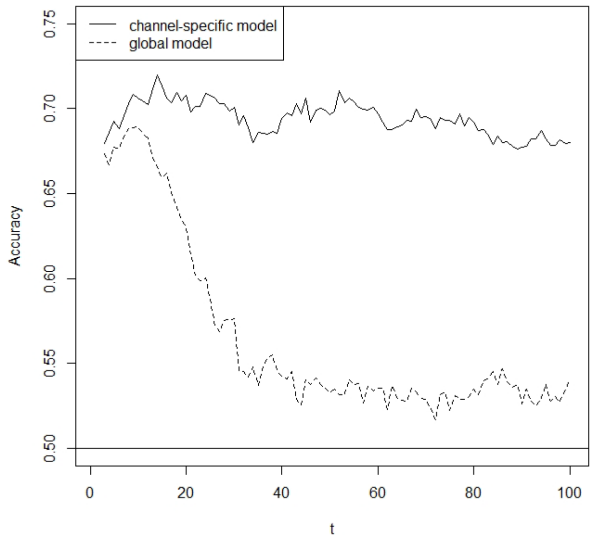

Figure 4. Accuracy of Random Forests over Time

Our findings offer evidence that we can indeed *learn* what personalities are associated with high-performing videos. Notably, accuracy could be improved by observing additional attributes of the video, including stochastic features such as view counts from prior days. However, to maintain our emphasis on content engineering, we focus on the relationship between the performance of a video and its time-invariant features, as only these features are manipulable during media creation.

Per Figure 4, channel-level accuracy remains consistently above the global accuracy. While the channel-level model peaks at 72% accuracy on day 16, the global accuracy peaks at 68% on day 12. This suggests there are both global and channel-level insights into the role of personality in video consumption, but that the channel-specific insights may be more informative. This is likely due to our prior argument that specific personalities may appeal to the target audience of each channel.

4.2 Economic Insights: Effect of Personality on Video Consumption

Building on our assessment of the predictive power of personality on video consumption, we assess the economic effects of personality classifications. To this end, we estimate Model 1 on both global and channel-specific forests to evaluate this economic impact. Estimating the global model produces results shown in Table 3[8]. Despite lower accuracy and robustness of global classifications, our estimates suggest that videos classified as high-performing based on their personality can expect to achieve 0.5% more views per day, relative to those with low-performing personalities. The positive coefficient on likelihood confirms the robustness of this result, and suggests that videos classified as high-performing with 10% greater confidence can expect to generate .028% more daily views.

Additionally, we find that videos with lower openness and lower conscientiousness may achieve more views. Our estimates indicate that a decrease of 10 percentile points in openness is associated with a 0.09% increase in daily views. Similarly, a 10 percentile decrease in conscientiousness is associated with a 0.13% increase in daily views. Finally, we find that one additional like or dislikes is associated with an increase in daily views by 0.002% and 0.006%, respectively. Because changes in likes and dislikes can be seen as a proxy for time-specific WOM associated with a video, we would expect to see this positive relationship. Furthermore, controlling for these time-specific factors allows us to control for exogenous shocks that may otherwise bias our estimates.

[8] Note that personality scores for both global and channel-specific models are normalized across all videos, and not just a specific subset of videos. In addition, multi-collinearity tests reveal no concerns over including all personality traits in one model – i.e., VIF is below 10 for all models.

Table 3. Effect of Global Performance Classification on Views

	Coefficient (SD)	Coefficient (SD)	Coefficient (SD)	Coefficient (SD)
Classification: $C_{ij,t}$	0.006*** (0.0003)	0.005*** (0.0003)	0.005*** (0.0003)	
Likelihood $L_{ij,t}$				0.028*** (0.0012)
Openness		-0.009*** (0.0026)	-0.005* (0.0026)	-0.005* (0.0026)
Conscientiousness		-0.013*** (0.0020)	-0.012*** (0.0019)	-0.012*** (0.0019)
Extroversion		0.003 (0.0017)	0.004 (0.0016)	0.004 (0.0017)
Agreeableness		0.002 (0.0014)	0.0002 (0.0014)	0.0001 (0.0015)
Neuroticism		0.002 (0.0024)	0.003 (0.0024)	0.003 (0.0024)
$\Delta Likes_{ij,t}$			0.00002*** (0.0000002)	0.00002*** (0.0000002)
$\Delta Dislikes_{ij,t}$			0.00006*** (0.000002)	0.00006*** (0.000002)
Intercept	0.017*** (0.0004)	0.029*** (0.0026)	0.022*** (0.0026)	0.011*** (0.0027)

*p < 0.1; *p < 0.05; ***p < 0.01

We observe in Table 4 that channel-specific personality classifications yield a one percent increase in daily views, having double the benefit of global classifications. Such a result is to be expected, given the greater accuracy and robustness of the channel-specific classifications. Supporting the robustness of this finding, our estimates suggest that videos classified as high-performing with 10% greater likelihood can expect to achieve 0.033% more daily views. Furthermore, we confirm our prior result on the negative effect of conscientiousness on video views, with our estimates suggesting that a 10 percentile decrease in conscientiousness would lead to a .17% increase in daily views.

Table 4. Effect of Channel-Specific Performance Classification on Views

	Coefficient (SD)	Coefficient (SD)	Coefficient (SD)	Coefficient (SD)
Classification: $C_{ij,t}$	0.010*** (0.001)	0.011*** (0.001)	0.010*** (0.001)	
Likelihood $L_{ij,t}$				0.033*** (0.002)
Openness		0.004 (0.11)	-0.001 (0.008)	0.003 (0.008)
Conscientiousness		-0.017** (0.008)	-0.013** (0.006)	-0.011* (0.006)
Extroversion		0.005 (0.007)	0.005 (0.005)	0.004 (0.005)
Agreeableness		0.006 (0.006)	0.0004 (0.005)	0.0002 (0.005)
Neuroticism		0.014 (0.010)	0.008 (0.008)	0.009 (0.008)
$\Delta Likes_{ij,t}$			0.00001*** (0.0000001)	0.00001*** (0.0000001)
$\Delta Dislikes_{ij,t}$			0.0002*** (0.0000003)	0.0002*** (0.0000004)
Intercept	0.035*** (0.002)	0.027*** (0.011)	0.030*** (0.001)	0.015* (0.009)

*p < 0.1; *p < 0.05; ***p < 0.01

A notable difference between the channel-specific and global model estimates is that while the global model suggests a negative effect of openness, the channel-level model does not. Given the greater accuracy and robustness of the channel model, this inconsistency may be due to the superior ability of the channel-level classifications to capture a preference for lower openness. At the same time, the notion that lower openness could be beneficial to many videos is supported by the high overall level of openness of YouTube videos (per Table 2), relative to the average openness of individuals. Because videos exhibiting high openness may not reflect the personality of their consumers, these viewers may be less likely to spread WOM about the video.

Given the aforementioned findings, we would like to illustrate the long-term, cumulative effects of our personality-based performance classifications. To do so, we apply the one percent increase in daily views suggested by our channel-level classifications across the lifespan of all videos in our dataset. We find that this daily increase in views increases cumulative views of videos by an average of 14.6% (SD = 3.2%). Hence, the daily increases in views associated with high-performing personalities may compound over time to generate meaningful long-run benefits. It should also be noted that content creators can reap these benefits across all videos from their channel, or a specific campaign, which could increase the effectiveness of their content marketing strategy.

5 Discussion

This research takes initial steps to evaluate the important yet underexplored relationship between content features and the consumption of digital media. Given growing emphasis on rich digital media, it is important to understand how the design of these media may influence their attractiveness. Through the lens of personality, we provide evidence of an association between the personality and the success of 6,440 videos from 80 YouTube channels. We find that personality can identify high-performing media with strong accuracy and robustness, and with meaningful benefits.

The implications of our work concern marketers, as well as other creators of rich digital media. First, these content creators should consider our global findings regarding what personalities tend to be associated with high-performing videos. By creating content that reflects a high-performing personality, they can create digital media that has a higher likelihood of achieving greater popularity. Second, content creators may want to obtain more detailed insights into what personalities are high-performing for their audience. We find that these channel-specific insights are superior in both accuracy and impact. This is likely due to the specificity of each channel's audience, which may not be a representative sample of the population, or even of the YouTube community. Thus, there may be specific personalities that appeal to these audiences.

Regarding specific personality traits, we find that videos of lower conscientiousness tend to achieve greater views. These videos may project a personality that is more spontaneous, disorderly, and fun. Such personalities are often associated with non-conformity and social differentiation (DeYoung et al. 2002). Due to individuals' desire for social differentiation and originality (Lemaine 1974), more consumers may be drawn to such videos. Because individuals enjoy differentiating themselves through their consumption preferences (Vandecasteele and Geuens 2010), these individuals may also be more prone to sharing WOM about these videos.
In addition to our insights into the role of personality, we introduce a novel methodology that can identify high-performing content features in general. This approach allows for the consideration of complex, multi-dimensional constructs such as visual features and tone, which are of increasing relevance due to advancements in content analysis techniques. Future research can consider

employing this approach to understand the role of other features for other forms of media.

References

Aaker JL (1997) Dimensions of Brand Personality. Journal of Marketing Research 34(3):347–356.

Adamopoulos P, Ghose A, Todri V (2018) The Impact of User Personality Traits on Word of Mouth: Text-Mining Social Media Platforms. Information Systems Research.

Baek TH, Morimoto M (2012) Stay Away From Me. Journal of Advertising 41(1):59–76.

Bampo M, Ewing MT, Mather DR, Stewart D, Wallace M (2008) The Effects of the Social Structure of Digital Networks on Viral Marketing Performance. Information Systems Research 19(3):273–290.

Barrick MR, Mount MK (1991a) The Big Five Personality Dimensions and Job Performance: A Meta-Analysis. Personnel Psychology 44(1):1–26.

Briley DA, Tucker-Drob EM (2014) Genetic and Environmental Continuity in Personality Development: A Meta-Analysis. Psychol Bull 140(5):1303–1331.

Byrne D (1997) An Overview (and Underview) of Research and Theory within the Attraction Paradigm. Journal of Social and Personal Relationships 14(3):417–431.

Corr PJ, Matthews G eds. (2009) The Cambridge handbook of personality psychology (Cambridge University Press, Cambridge, U.K. ; New York).

Costa PT, McCrae RR (1992) Normal Personality Assessment in Clinical Practice: The NEO Personality Inventory. Psychological Assessment 4(1):5–13.

Devaraj S, Easley RF, Crant JM (2008) Research Note—How Does Personality Matter? Relating the Five-Factor Model to Technology Acceptance and Use. Information Systems Research 19(1):93–105.

Devendorf SA, Highhouse S (2008) Applicant–employee similarity and attraction to an employer. Journal of Occupational and Organizational Psychology 81(4):607–617.

DeYoung CG, Peterson JB, Higgins DM (2002) Higher-order factors of the Big Five predict conformity: Are there neuroses of health? Personality and Individual Differences 33(4):533–552.

Doss SK, Carstens DS (2014) Big Five Personality Traits and Brand Evangelism. International Journal of Marketing Studies 6(3):13.

Fast LA, Funder DC (2008) Personality as manifest in word use: correlations with self-report, acquaintance report, and behavior. J Pers Soc Psychol 94(2):334–346.

Golbeck J, Robles C, Edmondson M, Turner K (2011) Predicting Personality from Twitter. 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing. 149–156.

Hirsh JB, Peterson JB (2009) Personality and language use in self-narratives. Journal of Research in Personality 43(3):524–527.

Järvinen J, Taiminen H (2016) Harnessing marketing automation for B2B content marketing. Industrial Marketing Management 54:164–175.

Kelly EL, Conley JJ (1987) Personality and compatibility: a prospective analysis of marital stability and marital satisfaction. J Pers Soc Psychol 52(1):27–40.

Lee K, Mahmud J, Chen J, Zhou M, Nichols J (2014) Who Will Retweet This?: Automatically Identifying and Engaging Strangers on Twitter to Spread Information. Proceedings of the 19th International Conference on Intelligent User Interfaces. IUI '14. (ACM, New York, NY, USA), 247–256.

Lemaine G (1974) Social differentiation and social originality. European Journal of Social Psychology 4(1):17–52.

Lichtenthal JD, Tellefsen T (2001) Toward a Theory of Business Buyer-Seller Similarity. Journal of Personal Selling & Sales Management 21(1):1–14.

Mahajan V, Peterson RA (1985) Models for innovation diffusion (Sage Publications, Beverly Hills).

McElroy JC, Hendrickson AR, Townsend AM, DeMarie SM (2007) Dispositional Factors in Internet Use: Personality versus Cognitive Style. MIS Quarterly 31(4):809–820.

Morry MM (2007) The attraction-similarity hypothesis among cross-sex friends: Relationship satisfaction, perceived similarities, and self-serving perceptions. Journal of Social and Personal Relationships 24(1):117–138.

Nass C, Lee KM (2001) Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. Journal of Experimental Psychology: Applied 7(3):171–181.

Norman WT (1963) Toward an adequate taxonomy of personality attributes: replicated factors structure in peer nomination personality ratings. J Abnorm Soc Psychol 66:574–583.

Pinto H, Almeida JM, Gonçalves MA (2013) Using Early View Patterns to Predict the Popularity of Youtube Videos. Proceedings of the Sixth ACM International Conference on Web Search and Data Mining. WSDM '13. (ACM, New York, NY, USA), 365–374.

Roccas S, Sagiv L, Schwartz SH, Knafo A (2002) The Big Five Personality Factors and Personal Values. Pers Soc Psychol Bull 28(6):789–801.

Salgado JF (2002) The Big Five Personality Dimensions and Counterproductive Behaviors. International Journal of Selection and Assessment 10(1–2):117–125.

Schwartz SH (1992) Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries. Zanna MP, ed. Advances in Experimental Social Psychology. (Academic Press), 1–65.

Susarla A, Oh JH, Tan Y (2011) Social Networks and the Diffusion of User-Generated Content: Evidence from YouTube. Information Systems Research 23(1):23–41.

Szabo G, Huberman BA (2010) Predicting the Popularity of Online Content. Commun. ACM 53(8):80–88.

Tucker CE (2014) Social Networks, Personalized Advertising, and Privacy Controls. Journal of Marketing Research 51(5):546–562.

Vandecasteele B, Geuens M (2010) Motivated Consumer Innovativeness: Concept, measurement, and validation. International Journal of Research in Marketing 27(4):308–318.